APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2008, p. 6720–6729 Vol. 74, No. 21
0099-2240/08/$08.00+0 doi:10.1128/AEM.00968-08

Hydrogenomics of the Extremely Thermophilic Bacterium *Caldicellulosiruptor saccharolyticus*▽†

Harmen J. G. van de Werken,[1]‡ Marcel R. A. Verhaart,[1]‡ Amy L. VanFossen,[2]‡ Karin Willquist,[3] Derrick L. Lewis,[2] Jason D. Nichols,[2] Heleen P. Goorissen,[1] Emmanuel F. Mongodin,[4] Karen E. Nelson,[5] Ed W. J. van Niel,[3] Alfons J. M. Stams,[1] Donald E. Ward,[6] Willem M. de Vos,[1] John van der Oost,[1] Robert M. Kelly,[2] and Servé W. M. Kengen[1]*

Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands[1]; Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695-7905[2]; Applied Microbiology, Lund Institute of Technology, Lund University, SE-221 00 Lund, Sweden[3]; Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland 21201[4]; The J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, Maryland 20850[5]; and Genencor International, a Danisco Company, 925 Page Mill Road, Palo Alto, California 94304[6]

Received 29 April 2008/Accepted 27 August 2008

***Caldicellulosiruptor saccharolyticus* is an extremely thermophilic, gram-positive anaerobe which ferments cellulose-, hemicellulose- and pectin-containing biomass to acetate, CO_2, and hydrogen. Its broad substrate range, high hydrogen-producing capacity, and ability to coutilize glucose and xylose make this bacterium an attractive candidate for microbial bioenergy production. Here, the complete genome sequence of *C. saccharolyticus*, consisting of a 2,970,275-bp circular chromosome encoding 2,679 predicted proteins, is described. Analysis of the genome revealed that *C. saccharolyticus* has an extensive polysaccharide-hydrolyzing capacity for cellulose, hemicellulose, pectin, and starch, coupled to a large number of ABC transporters for monomeric and oligomeric sugar uptake. The components of the Embden-Meyerhof and nonoxidative pentose phosphate pathways are all present; however, there is no evidence that an Entner-Doudoroff pathway is present. Catabolic pathways for a range of sugars, including rhamnose, fucose, arabinose, glucuronate, fructose, and galactose, were identified. These pathways lead to the production of NADH and reduced ferredoxin. NADH and reduced ferredoxin are subsequently used by two distinct hydrogenases to generate hydrogen. Whole-genome transcriptome analysis revealed that there is significant upregulation of the glycolytic pathway and an ABC-type sugar transporter during growth on glucose and xylose, indicating that *C. saccharolyticus* coferments these sugars unimpeded by glucose-based catabolite repression. The capacity to simultaneously process and utilize a range of carbohydrates associated with biomass feedstocks is a highly desirable feature of this lignocellulose-utilizing, biofuel-producing bacterium.**

Microbial hydrogen production from biomass has been recognized as an important source of renewable energy (15, 47). High-temperature microorganisms are well suited for production of biohydrogen from plant polysaccharides, as anaerobic fermentation is thermodynamically favored at elevated temperatures (17, 43). The extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* DSM 8903, a fermentative anaerobe initially isolated from wood in the flow of a thermal spring in New Zealand, first received attention because of its capacity to utilize cellulose at its optimal growth temperature, 70°C (37). Further work showed that *C. saccharolyticus* (i) can utilize a wide range of plant materials, including cellulose, hemicellulose, starch, and pectin, (ii) has a very high hydrogen yield (almost 4 mol of H_2 per mol of glucose) (14, 20, 48), and (iii) can ferment C_5 and C_6 sugars simultaneously. These features have led to the development of bioprocessing schemes based on *C. saccharolyticus*. For example, H_2 production is now being investigated using a two-step process in which H_2 and acetate are generated from biomass hydrolysates in one bioreactor and the acetate is fed to a second bioreactor and used by phototrophic organisms (*Rhodobacter* spp.) to produce additional H_2 in the presence of light (10). To provide a basis for full exploitation of the biohydrogen-producing capacity of *C. saccharolyticus*, the complete genome of this organism was sequenced and analyzed in conjunction with transcriptome information for the bacterium grown on glucose and xylose. The comparative and functional genomics approach that was employed here is referred to as "hydrogenomics." Insights arising from this effort revealed that *C. saccharolyticus* has the capacity to process and utilize a broad range of sugars and ultimately forms hydrogen from catabolism of these compounds.

* Corresponding author. Mailing address: Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands. Phone: 31-317-483737. Fax: 31-317-483829. E-mail: serve.kengen@wur.nl.

‡ H.J.G.V.D.W., M.R.A.V., and A.L.V. contributed equally to this work.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 5 September 2008.

MATERIALS AND METHODS

Cultivation and DNA isolation. *C. saccharolyticus* DSM 8903 (= ATCC 43494) was cultured overnight at 70°C on DSMZ 640 medium with glucose (50 mM) as the carbon and energy source. Cells were harvested and genomic DNA was isolated by the method of Pitcher et al. (35) using guanidinium thiocyanate.

TABLE 1. General features of the *C. saccharolyticus* genome

Characteristic	Value
Length of chromosome (bp)	2,970,275
G+C content (%)	35.3
Coding density (%)	86.3
Total no. of protein-encoding genes	2,679
Avg length of protein-encoding genes (bp)	958
Total no. of pseudogenes	92
Total no. of tRNA genes	46
Total no. of rRNA genes (no. of operons)	9 (3)
No. of CRISPR loci	9

Residual protein was removed in an additional purification step with sodium dodecyl sulfate and proteinase K, followed by chloroform-isoamyl alcohol extraction and isopropanol precipitation.

Genome sequencing and assembly. High-molecular-weight genomic DNA was provided to the U.S. Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) for cloning and shotgun sequencing. A combination of small (average insert sizes, 3 and 8 kb) and large (40 kb, fosmid) insert libraries were prepared and used for analysis, as indicated at http://www.jgi.doe.gov/.

Genome annotation and comparative analysis. Critica (2) and Glimmer (13) software programs were used for coding region detection and gene identification. TMHMM 2.0 (23) was used to predict transmembrane helices in translated sequences. SignalP v2.0b2 (33) was used to predict the presence and location of N-terminal signal peptides. All automatic gene and function predictions were manually checked using BLAST programs (1), InterPro (31), and the Integrated Microbial Genomes system (28) and corrected if necessary. Protein functions were checked by using the Carbohydrate-Active enzymes (CAZy) (12; http://www.cazy.org) classification. A comparative analysis was performed based on assignment and classification of Clusters of Orthologous Groups (COG) of proteins (44) by using the Integrated Microbial Genomes system.

Growth experiments and RNA isolation. *C. saccharolyticus* was subcultured overnight three times using the substrate of interest in modified DSMZ 640 medium before it was inoculated into a pH-controlled (pH 7) 1-liter fermentor containing 4 g substrate per liter. Cells were grown at 70°C until mid-logarithmic phase (optical density at 660 nm, 0.3 to 0.4) and were harvested by centrifugation and rapid cooling to 4°C and then stored at −80°C. Total RNA was isolated using a modified Trizol (Invitrogen) protocol in combination with an RNAeasy kit (Qiagen). Quality was determined with an Experion bioanalyzer (Bio-Rad), and cDNA was constructed with Superscript III reverse transcriptase (Invitrogen).

Whole-genome oligonucleotide DNA microarray design and construction. A DNA microarray was designed and constructed based on 2,695 protein-encoding sequences in the *C. saccharolyticus* genome obtained from the Department of Energy's Joint Genome Institute (http://genome.ornl.gov/microbial/csac). OligoArray 2.0 (40) was used to generate one 60-mer oligonucleotide probe sequence for each open reading frame. The probes were synthesized (Integrated DNA Technologies, Iowa), resuspended in 50% dimethyl sulfoxide, and printed on Ultragap microarray slides (Corning, New York) using a QArrayMini arrayer (Genetix, United Kingdom). Each probe was spotted five times on each array to fortify the statistical analysis.

Microarray hybridization. The cDNA samples were processed using a QIAquick purification kit (Qiagen, California), and they were eluted using phosphate buffer. The quantity and quality of the recovered cDNA samples were subsequently analyzed by using the ratio of absorbance at 260 to absorbance at 280 nm. Cyanine-3 and Cyanine-5 (Amersham, United Kingdom) dye labeling and sample hybridization were performed by following the instructions provided by TIGR (http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml), with minor adjustments to accommodate long-oligonucleotide platforms. Samples were hybridized using a four-slide loop (see Fig. S1 in the supplemental material).

Data collection and analysis. After incubation, slides were washed to remove nonspecifically bound material and scanned with a ScanArray Lite microarray scanner (Perkin Elmer, Massachusetts). Data acquisition and spot quantitation were performed with the ScanArray Express software. Once all the slides were quantitated, data from the loop were analyzed with JMP Genomics 3.0 (SAS, North Carolina), as described previously (36), using a mixed-effects analysis of variance model (49).

Nucleotide sequence and gene expression data accession numbers. The complete final genome assembly was released on 8 May 2007 and was deposited in the GenBank database under accession number CP000679 (http://genome.jgi-psf.org/finished_microbes/calsa/calsa.home.html). The gene expression data have been deposited in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/projects/geo/) under accession number GSE11153.

RESULTS

General features and comparative genomics of the genome of *C. saccharolyticus*. The genome of *C. saccharolyticus* DSM 8903 (= ATCC 43494) consists of one 2,970,275-bp circular chromosome, which has a G+C content of 35.3% (Table 1). The locations of the 2,679 predicted coding sequences on the two strands reflect the correlation between the direction of transcription and replication and show that the chromosome has two unequal replichores (Fig. 1). In addition, a GC skew analysis confirmed the huge size difference in the two replication arms, which might be attributed to a recent major inversion event. In addition to the protein-encoding genes, which were classified according to the COG system (Table 2), the chromosome harbors three rRNA operons and 46 tRNA genes with 41 different anticodons. These anticodons code for all 20 canonical amino acids. Like the chromosomes of many prokaryotes, the chromosome of *C. saccharolyticus* contains clustered regularly interspaced short palindromic repeats (CRISPR). CRISPR are DNA repeats that are separated by highly variable intervening sequences (spacers) and accompanied by CRISPR-associated (CAS) genes. The CRISPR and CAS protein products have been proposed to function as a defense mechanism against bacteriophages (4). With nine CRISPR loci and three different CAS genes, *C. saccharolyticus* is well equipped to fend off bacteriophages.

The complete genome sequence confirmed the phylogenetic position of *C. saccharolyticus* as a member of the class *Clos-*

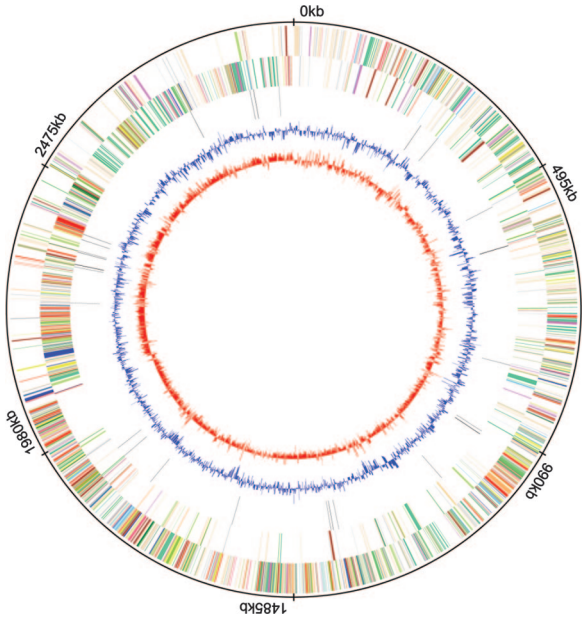

FIG. 1. Diagram of the *C. saccharolyticus* chromosome. From outside to inside the circles show (i) the genomic positions, (ii) the coding sequences on the positive negative strand, (iii) the coding sequences on the negative strand (the colors used for the coding sequences indicate the COG functional categories), (iv) the tRNA genes, (v) the G+C content, and (vi) the GC skew. The Microbial Genome Viewer was used to construct the circular chromosome wheel (21).

TABLE 2. Functional categories of predicted open reading frames in the genomes of hydrogen-producing organisms

COG category	COG functional category[a]	No. of proteins encoded in genomes[b]				
		C. saccharolyticus	*C. thermocellum*	*T. tengcongensis*	*T. maritima*	*P. furiosus*
Information storage and processing						
B	Chromatin structure and dynamics	2	1	2	1	4
L	Replication, recombination, and repair	222	252	149	89	109
A	RNA processing and modification	0	0	0	0	2
K	Transcription	134	174	141	82	80
J	Translation, ribosomal structure, and biogenesis	147	165	150	135	166
Cellular processes and signaling						
D	Cell cycle control, cell division, and chromosome partitioning	35	38	40	21	18
N	Cell motility	71	95	67	58	13
M	Cell wall, membrane, and envelope biogenesis	107	172	110	76	61
Z	Cytoskeleton	3	1	0	0	0
V	Defense mechanisms	48	40	42	27	29
U	Intracellular trafficking, secretion, and vesicular transport	42	60	46	39	20
O	Posttranslational modification, protein turnover, and chaperones	59	89	81	55	55
T	Signal transduction mechanisms	125	170	122	73	19
Metabolism						
E	Amino acid transport and metabolism	166	166	206	181	158
G	Carbohydrate transport and metabolism	213	144	160	166	93
H	Coenzyme transport and metabolism	101	102	67	62	89
C	Energy production and conversion	111	115	130	119	128
P	Inorganic ion transport and metabolism	73	91	96	113	95
I	Lipid transport and metabolism	34	46	55	31	25
F	Nucleotide transport and metabolism	56	62	62	54	52
Q	Secondary metabolite biosynthesis, transport, and catabolism	14	18	27	15	11
Poorly characterized						
S	Function unknown	177	187	173	137	189
R	General function prediction only	228	261	249	204	275
Not in COG		655	985	615	305	322

[a] Genes were classified with the Integrated Microbial Genomes system (28) using the COG functional classification for proteins (44).
[b] Number of protein-encoding genes in each category without pseudogenes.

tridia and revealed that *Thermoanaerobacter tengcongensis* (whose genome sequence has also been completed) is the closest relative (see Table S1 in the supplemental material). According to recently updated small-subunit rRNA databases, *C. saccharolyticus* and *T. tengcongensis* are assigned to the order *Thermoanaerobacterales*. The diversity in this order is, however, very great, and thus the order has limited biological significance. Reclassification is expected in the near future (26). The genome of *C. saccharolyticus* was compared to the genomes of two thermophilic relatives, *Clostridium thermocellum* and *T. tengcongensis* (3), as well as to the genomes of distantly related hyperthermophiles, including the bacterium *Thermotoga maritima* (32) and the archaeon *Pyrococcus furiosus* (39). These microorganisms have small to moderate-size genomes and also produce hydrogen when they grow on a range of carbohydrates (see Table S2 in the supplemental material). The genomic distributions of proteins in COG categories are comparable for this group of species (Table 2). However, one major difference is that both *C. thermocellum* and *C. saccharolyticus* have many more transposases and trans-

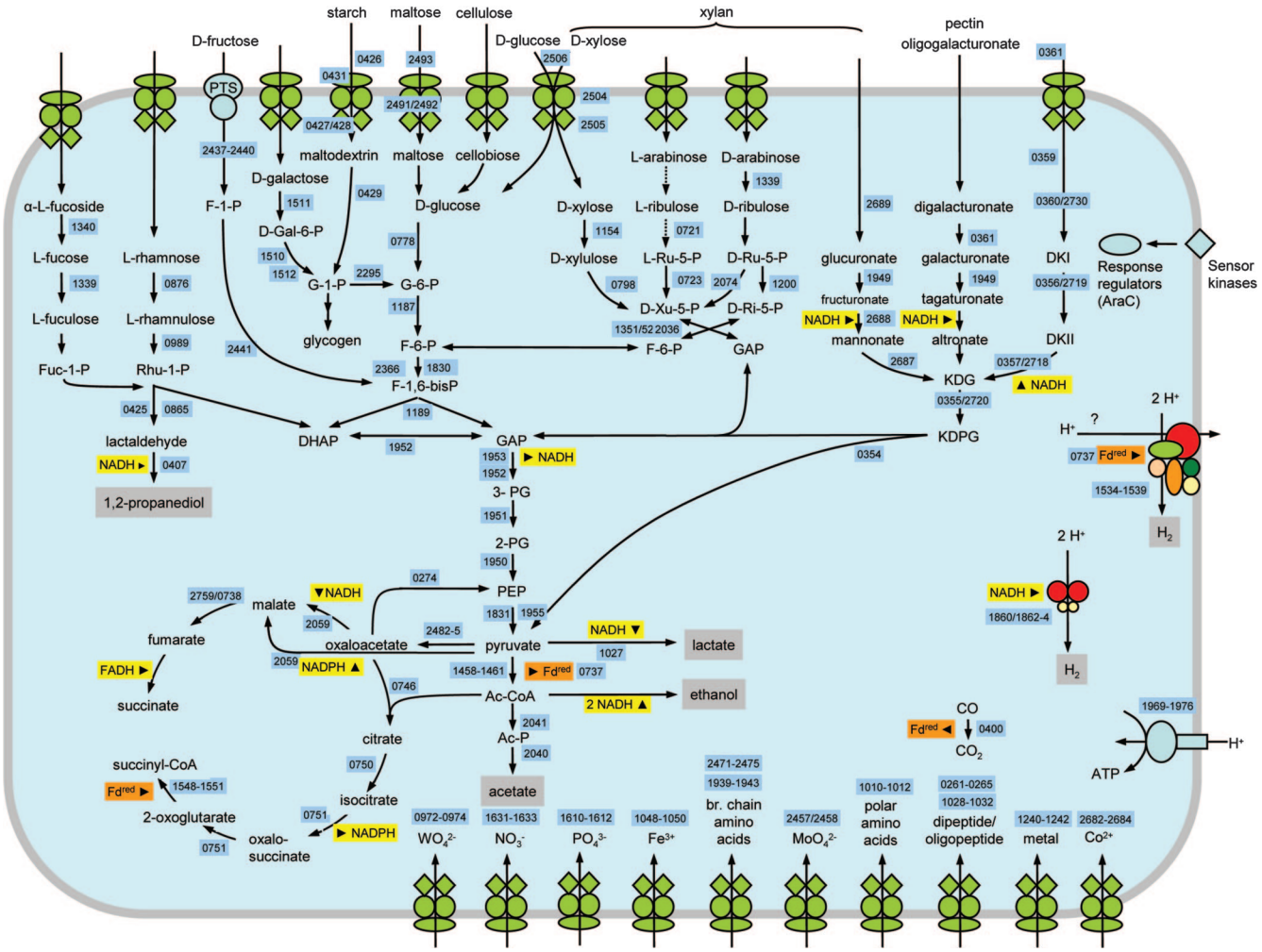

FIG. 2. Overview of the carbon metabolism and transport systems in *C. saccharolyticus*. The identities of the various ABC-type sugar transporters are not known. Secondary transport systems may be involved as well. Abbreviations: Rhu-1-P, rhamnulose-1-phosphate; F-1-P, fructose-1-phosphate; G-1-P, glucose-1-phosphate; G-6-P, glucose-6-phosphate; F-6-P, fructose-6-phosphate; F-1,6-bisP, fructose-1,6-bisphosphate; L-Ru-5-P, L-ribulose-5-phosphate; D-Ru-5-P, D-ribulose-5-phosphate; D-Xu-5-P, D-xylulose-5-phosphate; D-Ri-5-P, D-ribose-5-phosphate; DKI, 5-keto-4-deoxyuronate; DKII, 2,5-diketo-3-deoxygluconate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; KDPG, KDG phosphate; Fd^{red}, reduced ferredoxin; 3-PG, 3-phosphoglycerate; 2-PG, 2-phosphoglycerate; FADH, reduced flavin adenine dinucleotide; Ac-CoA, acetyl coenzyme A; Ac-P, acetyl phosphate; br., branched.

posase derivatives than the other species (99 and 93 such molecules, respectively). The genomes of *T. maritima*, *P. furiosus*, and *T. tengcongensis* encode 11, 17, and 43 transposases and transposase derivatives, respectively. Therefore, *C. saccharolyticus* and *C. thermocellum* have high numbers of proteins in the replication, recombination, and repair category (category L) (222 and 252 proteins, respectively). Furthermore, the *C. saccharolyticus* genome contains the largest number of carbohydrate transport and metabolism genes in this group. In fact, the *C. saccharolyticus* genome contains at least 177 ABC transporter genes, which is more than the 165 ABC transporter genes identified in *T. maritima* (11, 32). The *C. saccharolyticus* genome contains a reverse gyrase gene (Csac_1580), the product of which induces positive supercoiling of DNA (22). Reverse gyrase is considered a molecular marker of hyperthermophilicity and therefore distinguishes *C. saccharolyticus* from *C. thermocellum*, which lacks reverse gyrase. Despite the fact that *C. saccharolyticus* was described as a nonmotile organism, a set of flagellum structure, biogenesis, and chemotaxis genes was detected; it is not clear whether these genes are functional, since one of them is interrupted by a stop codon (pseudogene Csac_1277). *C. saccharolyticus* has a nitrogen fixation cluster (Csac_2461 to Csac_2466) and many sporulation genes; neither of these properties has been described previously for this bacterium.

Central carbon metabolism. *C. saccharolyticus* is able to metabolize a wide variety of carbohydrates, including the monosaccharides D-glucose, D-xylose, D-fructose, D-galactose, DL-arabinose, D-mannose, L-rhamnose, and L-fucose, as well as α- and β-linked di- and polysaccharides, including maltose, starch, pullulan, sucrose, trehalose, amorphous and microcrystalline cellulose, xylan, locust bean gum, and pectin (37). Once they are hydrolyzed, sugars are channeled to the central catabolic pathways (Fig. 2). Analysis of the genome sequence re-

vealed components of a complete Embden-Meyerhof (EM) pathway, including a ROK family glucokinase (Csac_0778), 6-phosphofructokinase, (Csac_2366 and Csac_1830), a bifunctional phosphoglucose/phosphomannose isomerase (Csac_1187), fructose-1,6-bisphosphate aldolase (Csac_1189), and pyruvate kinase (Csac_1831), as well as pyruvate-phosphate dikinase (PPDK) (Csac_1955). Also, a *gapA* operon was found, consisting of genes encoding glyceraldehyde-3-phosphate dehydrogenase (Csac_1953), the fusion protein phosphoglycerate kinase/triose-phosphate isomerase (Csac_1952), phosphoglycerate mutase (Csac_1951), and enolase (Csac_1950) (Fig. 2). However, the oxidative branch of the pentose phosphate pathway (PPP) and the Entner-Doudoroff pathway were not detected, which is consistent with previous findings obtained using ^{13}C nuclear magnetic resonance (14). The absence of the oxidative branch of the PPP, however, raises questions about how NADPH is produced for biosynthesis. The only other obvious NADPH-producing reaction is the isocitrate dehydrogenase (Csac_0751) reaction. However, based on sequence homology, the isocitrate dehydrogenase is likely to produce NADH instead of NADPH. Also, no obvious homolog of an NADPH-producing glyceraldehyde-3-phosphate dehydrogenase could be identified, as has been reported previously for *Streptococcus* species and some clostridia (8). NADPH can also be synthesized from NADH by a transhydrogenase, which is either membrane bound or soluble. In the genome there are no genes encoding orthologs of subunits of the membrane-bound type, but several homologous genes coding for soluble flavin adenine dinucleotide-dependent pyridine nucleotide-disulfide oxidoreductases are present (Csac_0759, Csac_2199, and Csac_0402). The exact physiological roles of these enzymes have not been determined. Furthermore, no ferredoxin:NADPH reductase homolog is present, although such activity has been found in some *Thermoanaerobacter* spp. (19).

Xylose, a major constituent of hemicellulose, is funneled by a putative xylose isomerase (Csac_1154) and a xylulokinase (Csac_0798) into the nonoxidative branch of the PPP. The nonoxidative branch of the PPP uses ribulose phosphate 3-epimerase (Csac_2074), ribose-5-phosphate isomerase (Csac_1200), the N-terminal (Csac_1351) and C-terminal (Csac_1352) transketolases, and transaldolase (Csac_2036) to produce the EM intermediates fructose-6-phosphate and glyceraldehyde-3-phosphate. Galactose also enters the EM via the Leloir pathway, which includes galactokinase (Csac_1511), galactose-1-phosphate uridylyltransferase (Csac_1510), UDP-glucose 4-epimerase (Csac_1512), and phosphoglucomutase (Csac_2295). Strikingly, genes encoding the established types of fructose bisphosphatase (classes I to IV [41]) are not evident in the *C. saccharolyticus* genome. Since fructose bisphosphatase is an essential enzyme in gluconeogenesis, *C. saccharolyticus* presumably uses a novel phosphatase. Moreover, a gene encoding the gluconeogenic phosphoenolpyruvate (PEP) synthetase is also not present, although conversion of pyruvate to PEP could occur via the reversible PPDK (Csac_1955) or via oxaloacetate.

Pyruvate, the end product of the EM pathway, is decarboxylated to acetyl coenzyme A (acetyl-CoA) by pyruvate:ferredoxin oxidoreductase (POR). *C. saccharolyticus* contains three 2-oxoacid:ferredoxin oxidoreductase enzyme complexes (Csac_2248 and Csac_2249, Csac_1458 to Csac_1461, and Csac_1548 to Csac_1551). According to transcriptional response data (see below), the true POR is probably encoded by the Csac_1458 to Csac_1461 genes. Acetyl-CoA is used to generate acetate and ATP (Csac_2040 and Csac_2041), or it enters the tricarboxylic acid (TCA) cycle for biosynthetic purposes. The TCA cycle in *C. saccharolyticus* is incomplete; it has an oxidative branch to succinyl-CoA catalyzed by a citrate (Re) synthase (Csac_0746), aconitate hydratase (Csac_0750), isocitrate dehydrogenase (Csac_0751), and the 2-oxoglutarate:ferredoxin oxidoreductase complex (Csac_1548 to Csac_1551). In the reductive direction, only orthologs of the subunits of fumarate hydratase were detected with a high level of confidence (Csac_2759/Csac_0738). Malate dehydrogenase (oxaloacetate decarboxylating) (Csac_2059) may be used to generate malate directly from pyruvate instead of from oxaloacetate. Fumarate reductase, however, could not be identified, which is in agreement with the absence of this enzyme in related clostridia. In addition to the activity of malate dehydrogenase, TCA metabolites could be replenished by a putative sodium pump oxaloacetate decarboxylase enzyme complex (Csac_2482 to Csac_2485).

Polysaccharide-degrading enzymes. The capacity of *C. saccharolyticus* to hydrolyze a broad range of polysaccharides prior to fermentation differentiates this bacterium from many thermophilic anaerobes. Indeed, the genome of *C. saccharolyticus* encodes a wide range of enzymes active with carbohydrates (see Table S3 in the supplemental material). The genes encoding these carbohydrate-utilizing enzymes are often clustered on the chromosome, and the enzymes can be assigned to substrate-specific catabolic pathways for cellulose, hemicellulose, and, to a lesser extent, starch and pectin. α-1,4-Glucan polymers, for instance, can be transported into the cell using the maltodextrin ABC transport system proteins (Csac_0427 to Csac_0428/Csac_0431). An intracellular α-amylase (Csac_0426) and a 1,4-α-glucan phosphorylase (Csac_0429) further degrade the intracellular maltodextrins, releasing glucose-1-phosphate. Remarkably, a transcriptional regulator of the LacI family (Csac_0430) is also in this maltodextrin cluster and therefore is a good candidate for controlling expression of this maltodextrin-degrading pathway at the transcriptional level. In addition, a GCAAACGTTTGC consensus sequence was found in sequences upstream of genes in this transport cluster and genes encoding several starch-degrading enzymes, such as an α-amylase precursor (Csac_0408), an oligo-1,6-glucosidase (Csac_2428), a pullulanase (Csac_0689), a 4-α-glucanotransferase (Csac_0203), and a putative glucan 1,4-α-glucosidase (Csac_0130). The consensus sequence resembles the binding site (CGCAAACGTTTGCGT) of the maltose/maltodextrin transcriptional repressor MalR from the gram-positive organism *Streptococcus pneumoniae* (34). Besides this putative starch-degrading regulon, *C. saccharolyticus* has genes encoding a glycogen metabolic cluster (Csac_0780 to Csac_0784), a maltose ABC transport system (Csac_2491 to Csac_2493), and a second pullulanase (Csac_0671). Thus, *C. saccharolyticus* is well equipped for starch utilization.

An important feature of *C. saccharolyticus* is its ability to produce H_2 not only from α-linked polymers but also from complex β-linked glycans, such as cellulose, hemicellulose, laminarin, and galactomannan. Growth on cellulosic substrates is rare among (hyper)thermophilic microorganisms. *C. saccharolyticus* does not metabolize cellulose by means of a cellulosome (45). For example, genes encoding the typical molecular

components of a cellulosome (i.e., dockerin domains and scaffolding proteins) were not identified in the genome. Nevertheless, a gene cluster (Csac_1076 to Csac_1081) containing genes encoding cellulase precursors is present. The cellulases are potentially capable of degrading cellulose, a plant polysaccharide (6) (see Table S3 in the supplemental material). Moreover, the product of another gene cluster (Csac_1089 to Csac_1091) and an extracellular cellulase (Csac_0678) may assist in completely hydrolyzing cellulose to glucose. A remarkable aspect of the β-glycanases in *Caldicellulosiruptor* species is their bifunctional domain architecture consisting of central cellulose binding domains which are bordered by distinct catalytic domains (6). The *C. saccharolyticus* genome sequence confirmed the presence of genes encoding these multidomain hydrolases (Csac_1076 to Csac_1079 and Csac_2411), which comprise cellulase, mannanase, xylanase and arabinofuranosidase domains (6, 27, 30, 45). No genes encoding additional bifunctional β-glycanases were found in the genome. The bifunctional Csac_2411 gene is part of another large gene cluster (Csac_2404 to Csac_2411) coding for enzymes involved in hydrolysis of the plant polysaccharide xylan (hemicellulose). The genes encoding these mostly extracellular enzymes might be coexpressed with genes in a smaller putative xylan-utilizing cluster (Csac_0203 to Csac_0205). The genes in the latter cluster were not significantly upregulated during growth on xylose, in contrast to the genes in the former cluster. Furthermore, putative genes that encode enzymes that degrade galactomannan (Csac_0663 and Csac_0664), galactoarabinan (Csac_1560 to Csac_1562), and laminarin (Csac_2548) were identified.

The plant cell wall component pectin consists of an α-1,4-linked D-galacturonic acid backbone, sometimes with L-rhamnose interspersed, and side chains consisting of monosaccharides, such as D-galactose, D-xylose, and L-arabinose (38). Degradation of the main pectin component, D-galacturonate, requires a galacturonate isomerase, a tagaturonate reductase, and an altronate dehydratase to form 2-keto-3-deoxygluconate (KDG). Galacturonate isomerization may be catalyzed by glucuronate isomerase (Csac_1949). However, genes encoding tagaturonate reductase and altronate dehydratase were not detected in the genome of *C. saccharolyticus*. Apparently, novel enzymes or a novel pathway is responsible for the degradation of galacturonate. In contrast, a gene cluster encoding enzymes involved in the conversion of glucuronic acid to KDG (Csac_2686 to Csac_2689) was identified, and it includes genes encoding fructuronate reductase, mannonate dehydratase, a putative β-galactosidase/β-glucuronidase, and an α-glucuronidase. Glucuronic acid is a common substituent of xylan. Enzymes involved in the conversion of KDG to pyruvate and glyceraldehyde-3-phosphate, including KDG kinase (Csac_0355 or Csac_2720) and KDG-6-phosphate aldolase (Csac_0354), are present as well. The genes encoding enzymes involved in the last two steps are clustered with genes (Csac_0356 and Csac_0357, as well as Csac_2718 and Csac_2719) that encode enzymes involved in the conversion of 5-keto-4-deoxyuronate, an unsaturated cleavage product of pectate, to KDG. The genes encoding enzymes that are able to hydrolyze the pectate backbone and the side chains (e.g., unsaturated rhamnogalacturonyl hydrolase [Csac_0360], galacturan 1,4-α-galacturonidase [Csac_0361], β-galactosidase [Csac_0362], and a glycoside hydrolase with unknown substrate specificity [Csac_0363]) are close to the genes encoding the KDG metabolic enzymes as well. However, neither a pectate lyase gene nor a methylesterase gene could be definitively identified in the genome, although Csac_2721 and Csac_2728 might be candidates for a pectate lyase gene based on distant homology to known lyase genes.

C. saccharolyticus is also able to grow on L-rhamnose and on L-fucose, producing 1,2-propanediol as the end product (unpublished data). A putative rhamnose catabolic pathway was identified, in which generation of dihydroxyacetone phosphate and 1,2-propanediol is catalyzed by an L-rhamnose isomerase (Csac_0876), a putative L-rhamnulokinase (Csac_0989), an L-rhamnulose-1-phosphate aldolase (Csac_0865), and a putative lactaldehyde reductase (Csac_0407). Fucose can be processed by a similar pathway, using the lactaldehyde reductase mentioned above and as-yet-unidentified versions of L-fuculokinase, a bifunctional L-fucose isomerase/D-arabinose isomerase (Csac_1339), and fuculose-1-phosphate aldolase (Csac_0425).

Fermentation products. Reducing equivalents are produced at the level of NAD and ferredoxin (Csac_0737). Since *C. saccharolyticus* can produce almost 4 mol of H_2 per mol of glucose (14), both NADH and reduced ferredoxin should ultimately be able to transfer their reducing equivalents to protons to form hydrogen. In the genome, two hydrogenase gene clusters were identified, which are very similar to the two related clusters in *T. tengcongensis* (42). The first cluster includes genes encoding subunits of an Ni-Fe hydrogenase (EchA to EchF) (Csac_1534 to Csac_1539) and various genes required for maturation of the hydrogenase complex (HypA to HypF) (Csac_1540 to Csac_1545). In *T. tengcongensis*, the Ni-Fe hydrogenase is ferredoxin dependent and membrane bound and may act as a proton pump to generate a proton motive force. The second cluster (Csac_1860 to Csac_1864) codes for an Fe-only hydrogenase (HydA to HydD), which is NAD dependent and located in the cytoplasm, similar to the situation in *T. tengcongensis* (42). Hydrogenases that form H_2 directly from NADH are unusual and make an NAD:ferredoxin oxidoreductase (Nfo) redundant. Nfos (also known as Rnfs) are membrane-bound multisubunit complexes that use or create an Na^+ gradient coupled to the transfer of reducing equivalents between NADH and ferredoxin (7). An Nfo gene cluster has been identified in the genomes of *C. thermocellum*, *T. maritima*, and *Thermotoga ethanolicus*, but not in the genomes of *T. tengcongensis* and *C. saccharolyticus*. The absence of an Nfo in *C. saccharolyticus* also implies either that in the presence of elevated levels of H_2 reduced ferredoxin may not be used to produce NADH or that a novel type of enzyme (complex) performs this reaction. Altogether, the information available suggests that *C. saccharolyticus* is able to produce hydrogen from ferredoxin but can also produce hydrogen directly from NADH. Production of hydrogen seems to be preferable, because under these conditions all pyruvate is converted to acetate (and CO_2), which is coupled to the synthesis of ATP.

When the hydrogen partial pressure is too high, formation of hydrogen from NADH is no longer thermodynamically favorable. In such a case, NADH is oxidized through the formation of lactate or ethanol. A gene for a lactate dehydrogenase was identified (Csac_1027), but genes for acetaldehyde dehydrogenase and alcohol dehydrogenase were not obvious. In *T. teng-*

congensis and *T. ethanolicus*, ethanol formation is NADPH dependent and catalyzed by a bifunctional alcohol dehydrogenase/acetyl-CoA thioesterase; this enzyme also has a homolog in *C. saccharolyticus* (Csac_0395).

A third small hydrogenase-like cluster was detected in the *C. saccharolyticus* genome and was composed of four genes encoding two NADH-binding proteins (Csac_0619 and Csac_0620), a molybdopterin oxidoreductase containing NAD and 4Fe-4S binding regions (Csac_0621), and an iron-containing alcohol dehydrogenase (Csac_0622). The function of this cluster is not known.

Transport systems. As mentioned above, a number of genes encoding ABC transporters have been found in the *C. saccharolyticus* genome, including genes encoding the previously described carbohydrate-specific maltodextrin ABC transport system (Csac_0427, Csac_0428, and Csac_0431) and the maltose ABC transport system (Csac_2491 to Csac_2493). As observed for certain *T. maritima* maltose transporters, both sets of transport proteins lack ATP-binding subunits. In many bacteria, the intracellular ATPase used in the system is not encoded in the same operon. The genes encoding both of these sets of ABC transporters are located downstream from the genes encoding a two-component system consisting of a sensor histidine protein kinase and a response regulator. In *C. saccharolyticus* ~50% of the genes encoding ABC carbohydrate transport systems are located near the genes encoding the two-component system on the chromosome.

Comparative analysis of *C. saccharolyticus* sugar binding proteins (SBPs) revealed that about two-thirds of them belong to the COG1653 category. This category includes the CUT1 subfamily members TM0432, TM0595, and TM1855, which transport a variety of di- and oligosaccharides, such as maltose. In genomes, the genes encoding more than one-half of COG1653 members are close to genes encoding glycoside hydrolases, supporting the identification of these proteins as ABC transporters involved in carbohydrate utilization. Putative SBPs encoded by Csac_0242, Csac_0391, Csac_2326, and Csac_2507 belong to COG1879. Csac_2506 and Csac_2510 are associated with the xylose transport-specific COG4213 category. As in *T. maritima*, a few putative SBPs (Csac_0261 and Csac_4166) are annotated as peptide transporters, although their actual functions are unknown. Components of phosphotransferase systems have been identified in *C. saccharolyticus* (although only one set of carbohydrate-specific EII has been identified), as have a few putative members of the major facilitator superfamily (Csac_0685, Csac_0786, Csac_1100, Csac1170, and Csac_2298). However, it is likely that carbohydrate utilization proceeds mainly through ABC transporters.

Transcriptional regulation. The ability of *C. saccharolyticus* to utilize many different carbohydrates suggests that there is tight regulation in the pathways. The genes encoding many enzymes involved in carbohydrate utilization pathways appear to be regulated at the transcriptional level. Apart from the RNA polymerase core enzyme subunits (Csac_2259, Csac_2085, Csac_0951, and Csac_0952), this gram-positive species has 12 different σ factors for construction of the RNA polymerase holoenzyme. In addition, multiple copies of genes encoding many sugar transcriptional regulators are present in the genome; for example, there are nine proteins belonging to the LacI family, six proteins belonging to the DeoR family, and eight proteins belonging to the GntR family, as well as 19 receiver proteins involved in a two-component system with a helix-turn-helix AraC domain. The genes encoding the latter proteins are always clustered with genes encoding sugar transporters and sugar hydrolytic enzymes.

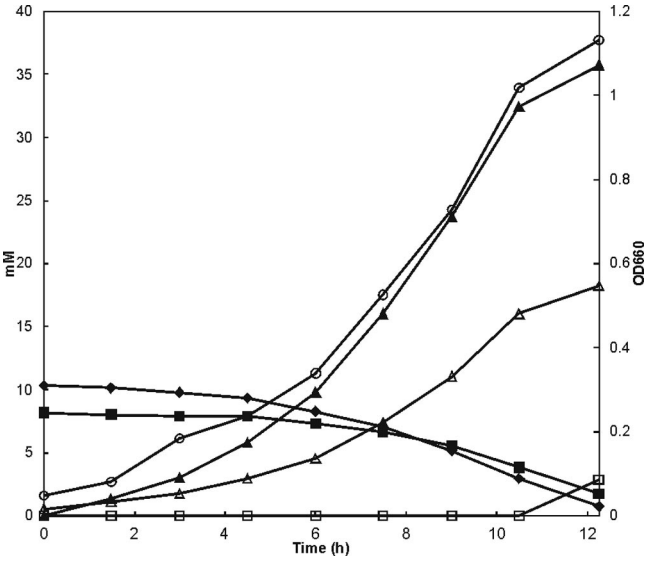

FIG. 3. Growth of *C. saccharolyticus* on a xylose-glucose mixture (1:1, wt/wt). ○, optical density at 660 nm (OD660); ▲, hydrogen; △, acetate; ■, glucose; ◆, xylose; □, lactate.

Carbon catabolite repression (CCR) by glucose was not observed in *C. saccharolyticus* (Fig. 3). Nevertheless, genes encoding some indicators of the presence of carbon control protein A (CcpA)-dependent CCR of gram-positive bacteria are present in the genome, including (i) a gene encoding a histidine-containing phosphocarrier (HPr) (Csac_2438) that is close to the gene encoding the only PEP-dependent phosphotransferase system, which is fructose specific; (ii) a gene encoding an HPr(Ser) kinase (Csac_1186); and (iii) genes encoding a catabolite repression HPr (CrH) (Csac_1163) and nine members of the CcpA-containing LacI family. Binding sites for a putative CcpA, the catabolite-responsive element (*cre*), could not be identified. The *Bacillus subtilis* consensus sequence WW TGNAARCGNWWWCAWW (29), for instance, was detected only twice, once in the upstream region of the gene encoding the α-amylase precursor mentioned above (Csac_0408), where it overlaps the putative MalR binding site, and once in the middle of the gene encoding fumarate hydratase subunit α (Csac_2759). Nevertheless, CCR probably occurs in *C. saccharolyticus*, although the metabolite that induces this repression is not known.

Besides global regulation through CCR, many local transcriptional regulators control the expression of carbohydrate metabolic pathways. Several orthologous transcriptional regulators were identified in *C. saccharolyticus*. The central glycolytic gene regulator (CggR) (Csac_1954), for instance, represses transcription of the *gapA* operon (25), while FruR (Csac_2442) controls the fructose operon (5). Based on the fact that many transcriptional regulator genes are close to their target operons, putative functions could be assigned to an α-linked glucan transcriptional regulator (Csac_0430), a regu-

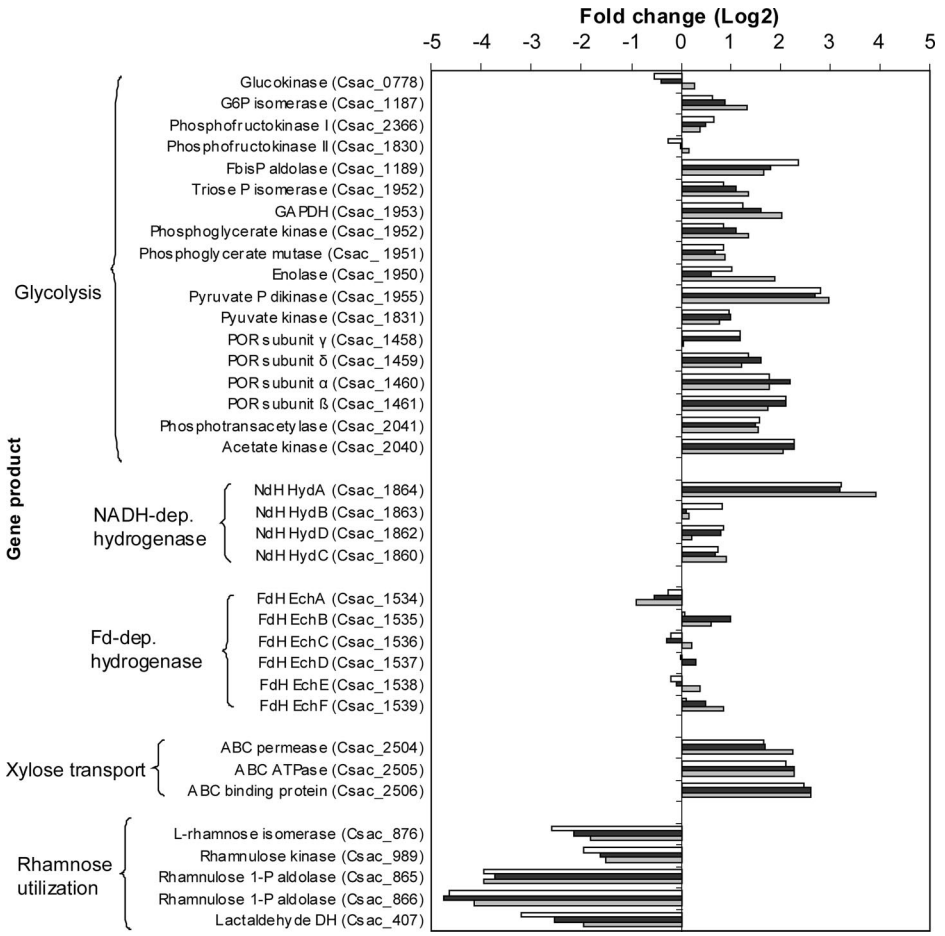

FIG. 4. Intensity ratios for transcript levels of selected genes that responded to growth on glucose (open bars), xylose (gray bars), or a mixture of glucose and xylose (black bars), compared to the results obtained with rhamnose. Ratios are expressed as $\log_2$ values. Abbreviations: GAPDH, glyceraldehyde-3-phosphate dehydrogenase; NdH, NAD-dependent hydrogenase; FdH, ferredoxin-dependent hydrogenase; dep., dependent; G6P, glucose-6-phosphate; FbisP, fructose bisphosphate; DH, dehydrogenase.

lator of the oxidative branch of the TCA cycle to oxoglutarate (Csac_0752), a repressor of L-arabinose metabolism (Csac_0722), and a putative response regulator receiver protein involved in glucuronate degradation (Csac_2690).

Transcriptome analysis of *C. saccharolyticus*. One of the features of *C. saccharolyticus* that is beneficial for hydrogen production is its ability to degrade cellulosic substrates as well as hemicellulose. Moreover, mixtures of glucose and xylose can be fermented simultaneously (Fig. 3), suggesting that classical CCR by glucose does not occur. To elucidate the central carbon metabolic pathways and their regulation, a transcriptome analysis was performed after growth on glucose, xylose, and a 1:1 mixture of these two substrates. L-Rhamnose, which was likely to be metabolized by another pathway, was used as a reference substrate. The transcriptional data clearly show that glucose, xylose, and the glucose-xylose mixture all triggered upregulation of genes in the EM pathway, compared to the results for rhamnose (Fig. 4; see Table S4 in the supplemental material). In particular, fructose bisphosphate aldolase, glyceraldehyde-3-phosphate dehydrogenase, PPDK, and POR were significantly stimulated. The ultimate acetate-forming acetate kinase was also highly upregulated. A catabolic role for PPDK is intriguing, since this enzyme normally is associated with gluconeogenesis (as it is in propionic acid bacteria and plants) and PEP is usually converted by pyruvate kinase. However, homologs of PPDK are also present in related clostridia and *Thermoanaerobacter* species.

It is worth noting that growth on glucose, growth on xylose, and growth on both sugars all triggered transcription of the genes encoding a xylose-specific ABC transport system (Csac_2504 to Csac_2506) (Fig. 4), suggesting that glucose and xylose are transported by the same uptake system. Moreover, none of the identified putative CCR genes (see below) were differentially transcribed, confirming that catabolite repression by glucose was not a factor.

The transcriptional response to growth on monosaccharides enabled identification of genes and groups of adjacent genes (gene clusters) that were specifically upregulated in response to either glucose or xylose. On glucose, several genes coding for α-glucan hydrolases responded. The most striking observation, however, was the upregulation of an entire gene cluster (Csac_1991 to Csac_2000) involved in purine synthesis, which was not observed with xylose. Upregulation of purine biosynthesis genes was also detected in the transcriptome of *Esche-*

richia coli growing on glucose compared to the results obtained with xylose (16). On xylose, several gene clusters required for xylan or xylose conversion were upregulated (Csac0692 to Csac_0696, Csac0240 to Csac_0242, and Csac2416 to Csac_2419). These clusters encode ABC transport systems, transcriptional regulators, and endoxylanases. In addition, genes specifically required for growth on rhamnose were highly upregulated during growth on rhamnose, indicating the pathway for utilization of this sugar.

DISCUSSION

C. saccharolyticus has been shown to be an excellent candidate for biohydrogen production (14, 20, 48). In contrast to mesophilic fermentative anaerobes, it produces almost no reduced end products, such as lactate or ethanol, and the amount of hydrogen produced approaches the "Thauer limit" of 4 mol of H_2/mol of glucose (17, 46). Moreover, *C. saccharolyticus* hydrolyzes various biomass-derived polymers, such as cellulose, hemicellulose, starch, and pectin and ferments corresponding sugar monomers, including glucose and xylose. The complete genome sequence of *C. saccharolyticus* provides new insights into the exceptional capacity of this bacterium to degrade a variety of plant polysaccharides and further reveals its high level of plasticity with many transposases, CRISPRs, and two uneven replication arms. A large number of sugar hydrolases and transferases were identified, and the number was greater than the number in the hyperthermophile *T. maritima* (9). Metabolic pathways for the degradation of residual components of cellulose, hemicellulose, starch, and pectin could be assigned. Reducing equivalents are produced as NADH or reduced ferredoxin, which are apparently used directly to produce hydrogen by a soluble NADH-dependent Fe-only hydrogenase and a membrane-bound ferredoxin-dependent Ni-Fe hydrogenase. The ability to produce hydrogen directly from NADH has not been observed for mesophilic anaerobes and may be responsible for the relatively high rates of hydrogen production by *C. saccharolyticus*. In mesophiles, reducing equivalents from NADH first have to be transferred to ferredoxin, which requires input of energy, either by a sodium gradient (7) or by coupling to an exergonic reaction (24). In extreme thermophiles, such as *C. saccharolyticus*, this is apparently not necessary.

The absence of catabolite repression by glucose is an important characteristic for biohydrogen producers since it allows them to process an array of biomass-derived substrates simultaneously. Glucose is generally known to repress the use of xylose by CCR (18), but this was not observed in *C. saccharolyticus*. Moreover, the transcriptome showed that the various components of a CCR system encoded in the genome (CcpA homologs, HPr, HPr kinase) were not differentially transcribed under the conditions examined, suggesting that this type of regulation is not triggered by glucose or xylose. There were no obvious differences in the transcriptome for central metabolic pathways during growth on glucose, xylose, or a mixture of glucose and xylose. The EM pathway was not affected by the hexose or pentose substrate, which is in contrast to the results of a transcriptome analysis of *E. coli* for growth on the same substrates (16). Remarkably, the same specific ABC transporter is upregulated with both substrates, which is also consistent with the nonpreferential behavior of *C. saccharolyticus* with these two monomeric sugars.

Detailed knowledge of the metabolic pathways leading to hydrogen production enables workers to identify key enzymes that may be targets for improving the H_2 yield by metabolic engineering. Currently, a genetic system for *C. saccharolyticus* is being developed, which initially will target the dehydrogenases involved in lactate and ethanol formation. Alternatively, genes could be introduced to form an oxidative branch of the PPP in order to obtain H_2 yields greater than 4 mol per mol of glucose (50). In any case, the *C. saccharolyticus* genome provides new insights into the metabolic features of a versatile biohydrogen producer, which could inspire efforts to optimize microbial bioenergy systems.

ACKNOWLEDGMENTS

We gratefully acknowledge the U.S. Department of Energy and the Joint Genome Institute for providing the genome sequence of *C. saccharolyticus*.

This research was supported by Earth and Life Sciences Foundation and Chemische Wetenschappen project 050.50.206 (subsidized by The Netherlands Organization for Scientific Research) and by the European Commission (HYVOLUTION project reference no. 19825). Microarray fabrication was financially supported in part by Danisco (Genencor International, Inc., Palo Alto, CA). R.M.K. acknowledges support from the U.S. NSF Biotechnology Program. A.L.V. and D.L.L. acknowledge support from a U.S. Department of Education GAANN Fellowship.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
3. **Bao, Q., Y. Tian, W. Li, Z. Xu, Z. Xuan, S. Hu, W. Dong, J. Yang, Y. Chen, Y. Xue, Y. Xu, X. Lai, L. Huang, X. Dong, Y. Ma, L. Ling, H. Tan, R. Chen, J. Wang, J. Yu, and H. Yang.** 2002. A complete sequence of the *T. tengcongensis* genome. Genome Res. **12:**689–700.
4. **Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709–1712.
5. **Barriere, C., M. Veiga-da-Cunha, N. Pons, E. Guedon, S. A. van Hijum, J. Kok, O. P. Kuipers, D. S. Ehrlich, and P. Renault.** 2005. Fructose utilization in *Lactococcus lactis* as a model for low-GC gram-positive bacteria: its regulator, signal, and DNA-binding site. J. Bacteriol. **187:**3752–3761.
6. **Bergquist, P. L., M. D. Gibbs, D. D. Morris, V. S. J. Te'o, D. J. Saul, and H. W. Morgan.** 1999. Molecular diversity of thermophilic cellulolytic and hemicellulolytic bacteria. FEMS Microbiol. Ecol. **28:**99–110.
7. **Boiangiu, C. D., E. Jayamani, D. Brugel, G. Herrmann, J. Kim, L. Forzi, R. Hedderich, I. Vgenopoulou, A. J. Pierik, J. Steuber, and W. Buckel.** 2005. Sodium ion pumps and hydrogen production in glutamate fermenting anaerobic bacteria. J. Mol. Microbiol. Biotechnol. **10:**105–119.
8. **Boyd, D. A., D. G. Cvitkovitch, and I. R. Hamilton.** 1995. Sequence, expression, and function of the gene for the nonphosphorylating, NADP-dependent glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus mutans*. J. Bacteriol. **177:**2622–2627.
9. **Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly.** 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J. Biol. Chem. **278:**7540–7552.
10. **Claassen, P. A. M., and T. de Vrije.** 2006. Non-thermal production of pure hydrogen from biomass: HYVOLUTION. Int. J. Hydrogen Energy **31:**1416–1423.
11. **Conners, S. B., C. I. Montero, D. A. Comfort, K. R. Shockley, M. R. Johnson, S. R. Chhabra, and R. M. Kelly.** 2005. An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **187:**7267–7282.
12. **Coutinho, P. M., and B. Henrissat.** 1999. Carbohydrate-active enzymes: an integrated database approach, p. 3–12. *In* H. J. Gilbert, G. Davies, B. Henrissat, and B. Svensson (ed.), Recent advances in carbohydrate bioengineering. The Royal Society of Chemistry Cambridge, United Kingdom.
13. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.

14. **de Vrije, T., A. E. Mars, M. A. Budde, M. H. Lai, C. Dijkema, P. de Waard, and P. A. Claassen.** 2007. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Appl. Microbiol. Biotechnol. **74**:1358–1367.
15. **European Commission.** 2002. Future needs and challenges for non-nuclear energy research in the European Union. Discussion paper. http://ec.europa.eu/research/energy/pdf/expert_workshops.pdf.
16. **Gonzalez, R., H. Tao, K. T. Shanmugam, S. W. York, and L. O. Ingram.** 2002. Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. Biotechnol. Prog. **18:**6–20.
17. **Hallenbeck, P. C.** 2005. Fundamentals of the fermentative production of hydrogen Water Sci. Technol. **52:**21–29.
18. **Hueck, C. J., and W. Hillen.** 1995. Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for the gram-positive bacteria? Mol. Microbiol. **15:**395–401.
19. **Hyun, H. H., G. J. Shen, and J. G. Zeikus.** 1985. Differential amylosaccharide metabolism of *Clostridium thermosulfurogenes* and *Clostridium thermohydrosulfuricum*. J. Bacteriol. **164:**1153–1161.
20. **Kadar, Z., T. de Vrije, G. E. van Noorden, M. A. Budde, Z. Szengyel, K. Reczey, and P. A. Claassen.** 2004. Yields from glucose, xylose, and paper sludge hydrolysate during hydrogen production by the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Appl. Biochem. Biotechnol. **113-116:** 497–508.
21. **Kerkhoven, R., F. H. van Enckevort, J. Boekhorst, D. Molenaar, and R. J. Siezen.** 2004. Visualization for genomics: the Microbial Genome Viewer. Bioinformatics **20:**1812–1814.
22. **Kikuchi, A., and K. Asai.** 1984. Reverse gyrase—a topoisomerase which introduces positive superhelical turns into DNA. Nature **309:**677–681.
23. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
24. **Li, F., J. Hinderberger, H. Seedorf, J. Zhang, W. Buckel, and R. K. Thauer.** 2008. Coupled ferredoxin- and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*. J. Bacteriol. **190:**843–850.
25. **Ludwig, H., G. Homuth, M. Schmalisch, F. M. Dyka, M. Hecker, and J. Stulke.** 2001. Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the *gapA* operon. Mol. Microbiol. **41:**409–422.
26. **Ludwig, W., K.-H. Schleifer, and W. B. Whitman.** Revised road map to the phylum Firmicutes. *In* P. De Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, S. K.-H., and W. B. Whitman (ed.), Bergey's manual of systematic bacteriology, 2nd ed., vol. 3. The Firmicutes, in press. Springer-Verlag, New York, NY. http://www.bergeys.org/outlines/Bergeys_Vol_3_Outline.pdf.
27. **Lüthi, E., D. R. Love, J. McAnulty, C. Wallace, P. A. Caughey, D. Saul, and P. L. Bergquist.** 1990. Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile "*Caldocellum saccharolyticum*". Appl. Environ. Microbiol. **56:**1017–1024.
28. **Markowitz, V. M., F. Korzeniewski, K. Palaniappan, E. Szeto, G. Werner, A. Padki, X. Zhao, I. Dubchak, P. Hugenholtz, I. Anderson, A. Lykidis, K. Mavromatis, N. Ivanova, and N. C. Kyrpides.** 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res. **34:**D344–348.
29. **Miwa, Y., A. Nakata, A. Ogiwara, M. Yamamoto, and Y. Fujita.** 2000. Evaluation and characterization of catabolite-responsive elements (*cre*) of *Bacillus subtilis*. Nucleic Acids Res. **28:**1206–1210.
30. **Morris, D. D., R. A. Reeves, M. D. Gibbs, D. J. Saul, and P. L. Bergquist.** 1995. Correction of the beta-mannanase domain of the *celC* pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. Appl. Environ. Microbiol. **61:**2262–2269.
31. **Mulder, N. J., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, V. Buillard, L. Cerutti, R. Copley, E. Courcelle, U. Das, L. Daugherty, M. Dibley, R. Finn, W. Fleischmann, J. Gough, D. Haft, N. Hulo, S. Hunter, D. Kahn, A. Kanapin, A. Kejariwal, A. Labarga, P. S. Langendijk-Genevaux, D. Lonsdale, R. Lopez, I. Letunic, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, A. N. Nikolskaya, S. Orchard, C. Orengo, R. Petryszak, J. D. Selengut, C. J. Sigrist, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats.** 2007. New developments in the InterPro database. Nucleic Acids Res. **35:**D224–D228.
32. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399:**323–329.
33. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
34. **Nieto, C., M. Espinosa, and A. Puyet.** 1997. The maltose/maltodextrin regulon of *Streptococcus pneumoniae*. Differential promoter regulation by the transcriptional repressor MalR. J. Biol. Chem. **272:**30860–30865.
35. **Pitcher, D. G., N. A. Saunders, and R. J. Owen.** 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol. **8:**151–156.
36. **Pysz, M. A., D. E. Ward, K. R. Shockley, C. I. Montero, S. B. Conners, M. R. Johnson, and R. M. Kelly.** 2004. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8:**209–217.
37. **Rainey, F. A., A. M. Donnison, P. H. Janssen, D. Saul, A. Rodrigo, P. L. Bergquist, R. M. Daniel, E. Stackebrandt, and H. W. Morgan.** 1994. Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiol. Lett. **120:**263–266.
38. **Ridley, B. L., M. A. O'Neill, and D. Mohnen.** 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry **57**:929–967.
39. **Robb, F. T., D. L. Maeder, J. R. Brown, J. DiRuggiero, M. D. Stump, R. K. Yeh, R. B. Weiss, and D. M. Dunn.** 2001. Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. Methods Enzymol. **330:**134–157.
40. **Rouillard, J. M., M. Zuker, and E. Gulari.** 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res. **31:**3057–3062.
41. **Sato, T., H. Imanaka, N. Rashid, T. Fukui, H. Atomi, and T. Imanaka.** 2004. Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. J. Bacteriol. **186:**5799–5807.
42. **Soboh, B., D. Linder, and R. Hedderich.** 2004. A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis*. Microbiology **150:**2451–2463.
43. **Stams, A. J.** 1994. Metabolic interactions between anaerobic bacteria in methanogenic environments. Antonie van Leeuwenhoek **66:**271–294.
44. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
45. **Te'o, V. S., D. J. Saul, and P. L. Bergquist.** 1995. *celA*, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. Appl. Microbiol. Biotechnol. **43:**291–296.
46. **Thauer, R.** 1976. Limitation of microbial hydrogen formation via fermentation, p. 201–294. *In* H. G. Schlegel and J. Barnea (ed.), Microbial energy conversion. Erich Goltze KG, Göttingen, Germany.
47. **United States Department of Energy.** 2002. A national vision of America's transition to a hydrogen economy—to 2030 and beyond. United States Department of Energy, Washington, DC. http://www1.eere.energy.gov/hydrogenandfuelcells/pdfs/vision_doc.pdf.
48. **van Niel, E. W. J., M. A. W. Budde, G. G. de Haas, F. J. van der Wal, P. A. M. Claasen, and A. J. M. Stams.** 2002. Distinctive properties of high hydrogen producing extreme thermophiles, *Caldicellulosiruptor saccharolyticus* and *Thermotoga elfii*. Int. J. Hydrogen Energy **27:**1391–1398.
49. **Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules.** 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Comput. Biol. **8:**625–637.
50. **Zhang, Y. H., B. R. Evans, J. R. Mielenz, R. C. Hopkins, and M. W. Adams.** 2007. High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. PLoS ONE **2:**e456.